

Rec'd PCT/PTO 03 JUN 2005

PCT/EP 03/15182

BUNDE REPUBLIK DEUTSCHLAND

12 DEC 2003

10/537486

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



REC'D 08 JAN 2004	
WIPO	PCT

#2

**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung**

Aktenzeichen: 102 56 935.5

Anmeldetag: 5. Dezember 2002

Anmelder/Inhaber: Siemens Aktiengesellschaft, München/DE

Bezeichnung: Auswahl der Benutzersprache an einem
rein akustisch gesteuerten Telefon

IPC: H 04 M, G 01 L

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der
ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 9. Dezember 2003
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Ebert

Beschreibung

Auswahl der Benutzersprache an einem rein akustisch gesteuerten Telefon

5

Bei Kommunikations- und Informationsgeräten werden textuelle Informationen in der durch die Länderversion vorgegebenen Sprache dargestellt. Damit einhergehend gibt es für den Benutzer die Möglichkeit, die gewünschte Sprache als

10 Benutzersprache oder Bediensprache einzustellen. Wenn jetzt - aus welchem Grund auch immer - die Sprache der Benutzeroberfläche verstellt ist, steht der Benutzer vor dem Problem, die gewünschte Benutzersprache wieder einzustellen, ohne dass er zu dem betreffenden Menüeintrag oder
15 Steuerungszustand durch textuelle Rückmeldungen geführt werden kann.

Dieses Problem besteht allgemein und ist nicht auf grafische Benutzeroberfläche mit Tastatur- oder Mauseingabe beschränkt.

20 Im Gegenteil wird es in Zukunft mehr und mehr Endgeräte geben, die rein akustisch bedient werden. Auch bei Call-Centern, bei denen ebenfalls eine rein akustische Bedienung vorgenommen wird, stellt sich das Problem. Dabei erfolgt die Spracheingabe über Spracherkennung und die Sprachausgabe
25 entweder durch Abspielen vorgefertigter Sprachkonserven oder durch automatisierte Sprachsynthese in Form einer Text-zu-Sprache-Umwandlung.

Bei Geräten mit Bildschirm- bzw. Display- und Tastatureingabe
30 findet man zur Lösung des dargestellten Problems folgende Vorgehensweise: Im Allgemeinen gibt es die Möglichkeit, das Gerät auf die werkseitig vorhandene Spracheinstellung zurückzusetzen. Dies erfolgt meist durch eine bestimmte Tastenkombination. Ebenso gibt es Geräte, bei denen sich auf
35 einfache Weise ein Sprachmenü aktivieren lässt, indem der Benutzer die Zielsprache auswählen kann. Das sieht dann etwa wie folgt aus:

2

Deutsch
Français
English
Українець (Ukrainisch)
Românesc (Rumänisch)
...

Tabelle 1

In diesem Menü kann der Benutzer nun die gewünschte und damit einzustellende Benutzersprache auswählen. Eine solche
 5 Vorgehensweise ist für rein akustisch gesteuerte Geräte natürlich nicht möglich.

Davon ausgehend liegt der Erfindung die Aufgabe zugrunde, die Auswahl der Benutzersprache einer Vorrichtung mittels eines
 10 rein akustischen Verfahrens zu ermöglichen. Die Auswahlmöglichkeit soll auch gerade dann zur Verfügung stehen, wenn die Vorrichtung keine Unterstützung durch eine Anzeige leisten kann oder soll.

15 Diese Aufgabe wird durch die in den unabhängigen Ansprüchen angegebenen Erfindungen gelöst. Vorteilhafte Ausgestaltungen ergeben sich aus den Unteransprüchen. Durch die Erfindung kann die einzustellende Benutzersprache einer Vorrichtung einfach dadurch eingestellt werden, dass zur Auswahl der
 20 Benutzersprache einfach die einzustellende Benutzersprache gesprochen wird. Ein Engländer sagt also: "English", ein Deutscher sagt einfach "Deutsch", ein Franzose sagt "Français" und ein Ukrainer sagt "Ukrajins'kyj" (englische Transliteration von "ukrainisch" in polnischer Schrift).

25 Die Realisierung dieser Funktionalität in den Spracherkennungsmitteln der Vorrichtung ist keineswegs trivial, weshalb im Folgenden bevorzugte Möglichkeiten genauer dargestellt werden sollen.

Eine Möglichkeit besteht darin, einen Einzelworterkenner auf die zu erkennenden Bezeichnungen der einstellbaren Benutzersprachen zu trainieren. Da die hier verwendeten
5 Algorithmen zumeist auf bloßem Mustervergleich basieren, benötigt man für das Training eine ausreichende Anzahl von Sprachkonserven, in denen die Sprache (als Wort) von muttersprachlichen Sprechern zu der entsprechenden Sprache (als Code) aufgesprochen ist. Hierbei kann insbesondere ein
10 Dynamic-Time-Warp („DTW“)-Erkenner zu Einsatz kommen.

Sollte die Vorrichtung, etwa für andere Funktionalitäten, bereits über eine phonembasierte Spracherkennung verfügen, so ist es vorteilhaft, diese für das Einstellen der
15 Benutzersprache heranzuziehen. Dazu ergeben sich drei Möglichkeiten.

So kann in den Spracherkennungsmitteln ein multilinguales Hidden-Markov-Modell („HMM“) zur Anwendung kommen, das die
20 Phoneme aller Sprachen modelliert. Hierfür ist eine standardisierte Darstellung eines phonetischen Alphabets besonders vorteilhaft, etwa in der Form von SAMPA-Phonemen.

So überzeugend dieser Ansatz für die geschilderte Problemstellung ist, zeigen sich doch in der Praxis multilinguale Spracherkennungsmittel bezüglich der Erkennungsrate einer sprachspezifischen Modellierung unterlegen. Für die normale Spracherkennung in der Vorrichtung würde also noch ein weiteres akustisches Modell
30 benötigt, welches weiteren Speicherplatz verbräuchte.

Deshalb erweist sich eine andere Möglichkeit als vorteilhaft, bei der die zu den Bezeichnungen der einstellbaren Benutzersprachen gehörigen Phonemsequenzen aus den HMMs für
35 die unterschiedlichen Sprachen kombiniert sind. Hierbei muss allerdings bedacht werden, dass die Übereinstimmungsmaße, die die Spracherkennung für die in unterschiedlichen

Phoneminventaren modellierten Wörter liefert, nicht direkt miteinander vergleichbar sind. Dieses Problem kann umgangen werden, wenn im kombinierten HMM die Übereinstimmungsmaße für die Phonemsequenzen aus den unterschiedlichen erkennbaren Benutzersprachen skaliert sind.

Eine besonders geschickte Möglichkeit ergibt sich, wenn man statt eines multilingualen HMMs oder der Kombination von Phonemsequenzen mehrerer sprachspezifischer HMMs nur ein einziges sprachspezifisches bzw. länderspezifisches HMM verwendet und dabei die Bezeichnungen der fremden Benutzersprachen mit dem sprachspezifischen Phonemsatz modelliert. Zur Erläuterung dient das folgende Beispiel für die deutsche Sprache, das dem Menü in Tabelle 1 nachempfunden ist. Die Wortmodelle sind in "phonetischer" Schreibweise:

/ d e u t s h /
/ f r o n g s a e /
/ i n g l i s h /
/ u k r a i n s k i j /
/ r o m a n e s h t s h /

Tabelle 2

Hier entfällt die Notwendigkeit, ein multilinguales HMM zu verwenden oder Phonemsequenzen mit unterschiedlichen Phoneminventaren bei der Erkennung zu kombinieren.

Gemäß der einleitenden Problemstellung ist die Vorrichtung insbesondere ein mobiles Endgerät in Form eines Mobil- oder Schnurlostelefons, ein Headset oder der Server eines Call-Centers.

Bevorzugte Ausgestaltungen des erfindungsgemäßen Verfahrens ergeben sich analog zu den dargestellten bevorzugten Ausgestaltungen der erfindungsgemäßen Vorrichtung.

Weitere wesentliche Merkmale und Vorteile der Erfindung ergeben sich aus der Beschreibung eines Ausführungsbeispiels anhand der Zeichnung. Dabei zeigt:

5 Figur 1 den Ablauf bei der Einstellung der Benutzersprache.

Die Vorrichtung kann in Form eines schnurlosen Headsets verwirklicht sein, das ausschließlich über Sprache gesteuert wird. Dies kann beispielsweise ein Headset sein, das über
10 Bluetooth, Dect, GSM, UMTS, GAP oder einen anderen Übertragungsstandard mit oder ohne Kabel eine Verbindung zu einer Basis aufnimmt.

Das Headset weist eine On-/Off-Taste sowie eine so genannte
15 "P2T-Taste" ("push-to-talk") auf, mit der der Audiokanal für ein bestimmtes Zeitfenster auf die Spracherkennungsmittel geschaltet wird. Die Kommandosteuerung des Headsets beinhaltet das kurze Drücken der P2T-Taste, ein Quittieren des Tastendrucks durch einen kurzen Piep-Ton und das
20 anschließende Sprechen des gewünschten Kommandos, worauf die Vorrichtung entsprechend reagiert.

Beim ersten Einschalten der Vorrichtung (Schritt 1) bzw. nach einem Zurücksetzen der Vorrichtung (Schritt 2), was
beispielsweise durch ein längeres Drücken auf die P2T-Taste hervorgerufen wird, befindet sich der Benutzer zunächst in der Benutzersprachauswahl. Dies wird dem Benutzer durch ein akustisches Signal (Schritt 3) mitgeteilt, das beispielsweise
in einem längeren Piep-Ton oder einer mehrsprachigen
30 Aufforderung zum Einsprechen der einzustellenden Benutzersprache besteht.

Der Benutzer spricht nun die Bezeichnung der einzustellenden Sprache in der einzustellenden Sprache in die Vorrichtung ein
35 (Schritt 4). Die Spracherkennungsmittel der Vorrichtung erkennen nun die in der einzustellenden Benutzersprache gesprochene Bezeichnung der einzustellenden Benutzersprache,

soweit es sich bei der einzustellenden Benutzersprache um eine der mehreren einstellbaren Benutzersprachen der Vorrichtung handelt. Daraufhin stellen die Benutzerspracheneinstellungsmittel der Vorrichtung die

- 5 Benutzersprache der Vorrichtung auf die von den Spracherkennungsmitteln erkannte Benutzersprache ein, wodurch die Vorrichtung entsprechend initialisiert wird. Anschließend kann mit der Vorrichtung gearbeitet werden (Schritt 6), als ob sie normal eingeschaltet wurde (Schritt 5).

10

Bei der Korrektur von Spracherkennungs- und Bedienfehlern kann auf bewährte Mittel und Verfahren aus dem Stand der Technik zurückgegriffen werden.

- 15 Allen Ausführungsformen der Erfindung ist der herausragende Vorteil einer deutlichen Vereinfachung und Verkürzung der Bedienung der Vorrichtung gemeinsam. Weiterhin gibt es, sofern eine phonembasierte Erkennung verwendet wird, nicht die Notwendigkeit, Sprachkonserven in der Vorrichtung abzulegen. Die Tatsache, dass bereits phonembasierte
20 akustische Ressourcen in der Vorrichtung vorhanden sind, wird dabei optimal ausgenutzt.

Patentansprüche

1. Vorrichtung mit
 - Spracherkennungsmitteln zum Erkennen einer in einer einzustellenden Benutzersprache gesprochenen Bezeichnung der einzustellenden Benutzersprache der Vorrichtung, wobei mehrere einstellbare Benutzersprachen der Vorrichtung durch die Spracherkennungsmittel erkennbar sind, indem die Bezeichnung der jeweils einzustellenden Benutzersprache in der einzustellenden Benutzersprache gesprochen wird,
 - Benutzerspracheneinstellungsmitteln zum Einstellen der Benutzersprache der Vorrichtung auf die von den Spracherkennungsmitteln erkannte Benutzersprache.
2. Vorrichtung nach Anspruch 1, dadurch gekennzeichnet, dass die Spracherkennungsmittel einen Einzelworterkenner aufweisen.
3. Vorrichtung nach Anspruch 1, dadurch gekennzeichnet, dass die Spracherkennungsmittel einen phonembasierten Erkenner aufweisen.
4. Vorrichtung nach Anspruch 3, gekennzeichnet durch ein multilinguales Hidden-Markov-Modell, das von den Spracherkennungsmitteln verwendbar ist.
5. Vorrichtung nach Anspruch 3, gekennzeichnet durch ein kombiniertes Hidden-Markov-Modell, das Phonemsequenzen aus den mehreren erkennbaren Benutzersprachen enthält und von den Spracherkennungsmitteln verwendbar ist.
6. Vorrichtung nach Anspruch 5, dadurch gekennzeichnet,

dass im kombinierten Hidden-Markov-Modell die Übereinstimmungsmaße für die Phonemsequenzen aus den mehreren erkennbaren Bediensprachen skaliert sind.

5 7. Vorrichtung nach Anspruch 3,
gekennzeichnet durch
ein sprachspezifisches Hidden-Markov-Modell, bei dem die
Phoneme für die Bezeichnungen der mehreren erkennbaren
Benutzersprachen mit dem sprachspezifischen Phonemsatz des
10 sprachspezifischen Hidden-Markov-Modells modelliert sind.

8. Vorrichtung nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
dass die Vorrichtung ein mobiles Endgerät ist.

15 9. Vorrichtung nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
dass die Vorrichtung Mittel zur Ausgabe einer Aufforderung
zum Einsprechen der Bezeichnung der einzustellenden
20 Benutzersprache aufweist.

10. Verfahren zum Einstellen einer Benutzersprache einer
Vorrichtung,
- bei dem mehrere einstellbare Benutzersprachen durch eine
25 Spracherkennung erkennbar sind, wenn die Bezeichnung der
jeweils einzustellenden Benutzersprache in der
einzustellenden Benutzersprache gesprochen wird,
- bei dem die in der einzustellenden Benutzersprache
gesprochene Bezeichnung der einzustellenden Benutzersprache
30 erkannt wird,
- bei dem die erkannte einzustellende Benutzersprache als
Benutzersprache der Vorrichtung eingestellt wird.

Zusammenfassung

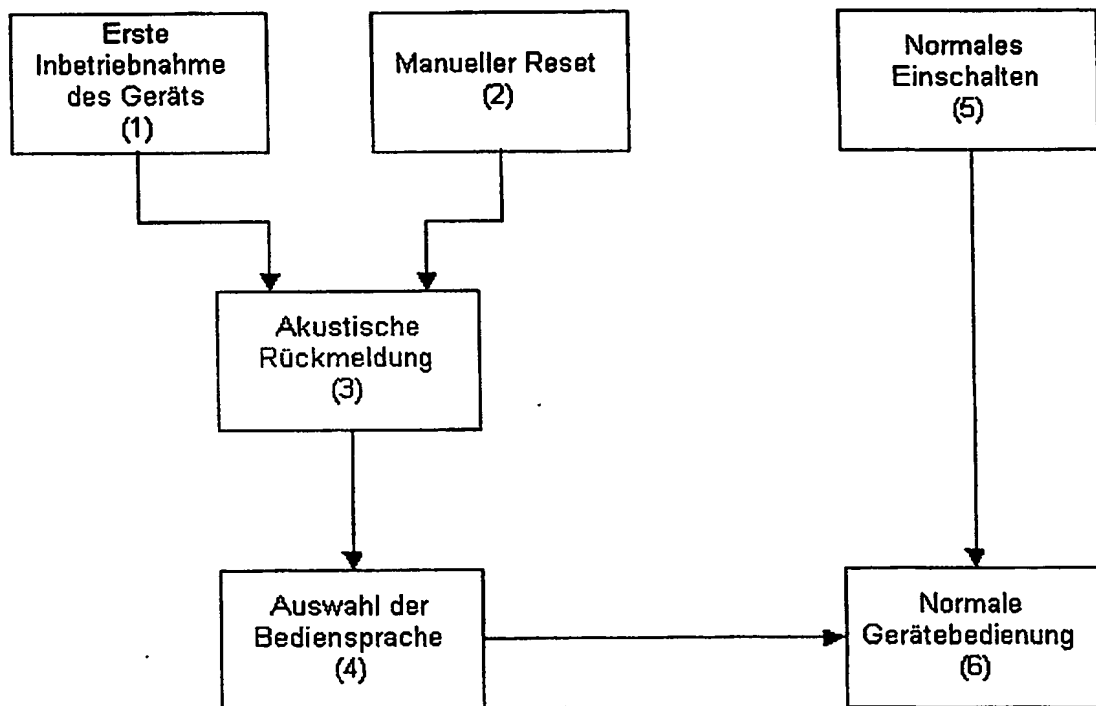
Auswahl der Benutzersprache an einem rein akustisch gesteuerten Telefon

5

Die Benutzersprache einer Vorrichtung kann durch Einsprechen der Bezeichnung der einzustellenden Benutzersprache in der einzustellenden Benutzersprache eingestellt werden.

10

Fig. 1



PCT Application

EP0313182

